

A19 - Nutrition Facts for US McDonald's Menu

TEAM: Max Frolov, Kaspar Saarem

TASK 1. Setting up

Link to repository: <https://github.com/kasparsaarem/McDonald-s>

TASK 2. Business understanding

- Identifying business goals

- Background

This project is not meant to benefit a business. This project is purely for us as individuals to get more information about the US McDonald's menu and find interesting facts from the data. McDonald's is one of the world's largest and famous fast food chains in the world and they provide a nutrition analysis of menu items to help customers balance their McDonald's meal with other foods they eat. Key highlight of McDonald's is that customers can view the complete nutritional information about the foods. So here we will analyse the various categories McDonald's offers, and its menu items and their nutritional values.

- Business goals

In this project our goal is to find out if McDonald's menu is actually as unhealthy as people claim it to be. We will find out answers to questions like how does average daily caloric need for a kid compare to calories in McDonald's Kids menu, How many calories and macro nutrients does McDonald's value meal contain on average, What is the least number of items you could order from the menu to meet one day's nutritional requirements, McDonald's food is thought to have high level of cholesterol and sodium which can be associated with increased risk in cardiovascular problems, our goal is to explore content of daily average % of cholesterol and sodium in McDonald's menu versus other nutrition. Find the healthiest combination of items from McDonald's menu to get daily nutritional requirements

- Business success criteria

Success criteria will be based on information found on for some of the questions. https://www.nutrition.org.uk/attachments/article/234/Nutrition%20Requirements_Revised%20Oct%202016.pdf .Like how does average daily caloric need for a kid compare to calories in McDonald's Kids menu if the calories from the average meal is higher than 50% of kids caloric need, or when seeing the average % of cholesterol and sodium of menu items we can compare them to % of daily values needed for a human and if the values stand out then the assumption is correct. For question finding the healthiest combination of items from McDonald's menu to get daily nutritional requirements will be compared to the ideal nutritional requirements.

- Assessing the situation

- Inventory of resources

- 1) https://www.kaggle.com/mcdonalds/nutrition-facts?fbclid=IwAR03NzFIE_3T49vU8O8SHldh8huGkfLxAXOp3P1cX9JKi1tXuEXC6TCymBs
- 2) https://www.nutrition.org.uk/attachments/article/234/Nutrition%20Requirements_Revised%20Oct%202016.pdf
- 3) <https://www.mcdonalds.com/us/en-us.html>

- Requirements, assumptions, and constraints

We have the legal access to all the resources we need. Finished work must contain the answers of all the data-mining goals and few healthy tips inferred from the information obtained during the process. The project should be finished before 16th of December.

- Risks and contingencies

Cause - the health of the team member has deteriorated badly because of COVID-19 and he is unable to contribute to the project - Solution: the other team member is ready to do more work. If the deadline is too close, additional time will be requested.

Cause: could not make the program do what we want it to do. Solution: we spend some time googling the suitable solution, in the extreme case, we seek help from fellow students or write to the tutor.

- Terminology

data

Values collected through record keeping or by polling, observing, or measuring, typically organized for analysis or decision making. More simply, data is facts, transactions and figures.

data format

Data items can exist in many formats such as text, integer and floating-point decimal. Data format refers to the form of the data in the database.

exploratory analysis

Looking at data to discover relationships not previously detected. Exploratory analysis tools typically assist the user in creating tables and graphical displays.

clustering

Clustering algorithms find groups of items that are similar. For example, clustering could be used by an insurance company to group customers according to income, age, types of policies purchased and prior claims experience. It divides a data set so that records with similar content are in the same group, and groups are as different as possible from each other. Since the categories are unspecified, this is sometimes referred to as unsupervised learning.

mean

The arithmetic average value of a collection of numeric data.

prevalence

The measure of how often the collection of items in an association occurs together as a percentage of all the transactions. For example, "In 2% of the purchases at the hardware store, both a pick and a shovel were bought."

range

The range of the data is the difference between the maximum value and the minimum value. Alternatively, range can include the minimum and maximum, as in "The value ranges from 2 to 8."

visualization

Visualization tools graphically display data to facilitate better understanding of its meaning. Graphical capabilities range from simple scatter plots to complex multi-dimensional representations.

- **Costs and benefits**

This project does not have any costs. The benefits highly exceed the costs. The biggest benefit is the knowledge of what is the “perfect” order a customer can make in McDonald’s. What order the customer should make to not feel guilty after the visit to McDonald’s, because his/her daily nutrient proportions are still balanced.

- **Data-mining**

- **Goals**

The purpose of data mining is to group data by each function and to extract from each category the closest results to people's daily needs. The uprooted data is collated and we try to find which foods and drinks are closest to the desired results.

Extract the closest result to the daily requirement of each nutrient, and from the results we try to put together the healthiest McDonald's meal.

We extract the average result for each nutrient and compare it with the needs of the average person to find out what makes fast food an unhealthy choice.

- **Success criteria**

The success criteria is to be able to make a model that takes test data from the dataset (McDonald’s menu) and train data from the ideal daily nutritional ratios.

TASK 3. Data understanding

- Gathering data

- Outline data requirements

We would need objects, int64 and float64 type data to analyze the numbers and compare them.

- Verify data availability

All of the data we use is publicly available.

- Selection criteria

Dataset - the dataset we use contains nutrition information of every item in the US McDonald's menu.

1. Item - name of the item in the menu.
2. Serving size - the weight of the meal.
3. Total Fat (% Daily value), Saturated Fat (% Daily Value), Cholesterol (% Daily Value), Carbohydrates (% Daily Value), Dietary Fiber (% Daily Value), Vitamin A (% Daily Value), Vitamin C (% Daily Value), Calcium (% Daily Value), Calcium (% Daily Value) - shows what percentage of the daily requirement of a given nutrient is covered by that food.

PDF file of the nutrition requirements - Human need for different nutrients
(Based on a study done in UK)

- Describing data

The dataset we use has 260 items and 24 features that are describing each item.

1. Category - where the item belongs.
2. Item - name of the meal/food.
3. Serving Size - the weight of the meal. (oz and grams or carton and ml)
4. Calories - the amount of energy we get from that food.
5. Calories from Fat - how much energy we get from fat.
6. Total fat - the amount of fat in that food. (grams)
7. Total Fat (% of Daily Value) - percentage of the daily requirement of fat covered by that food.

8. Saturated Fat - in the amount of fat, how many grams of that amount are covered by saturated fat.
9. Saturated Fat (% Daily Value) - percentage of the daily requirement of saturated fat covered by that food.
10. Trans Fat - the amount of trans fat in that food. (grams)
11. Cholesterol - the amount of cholesterol in that food. (mg)
12. Cholesterol (% Daily Value) - percentage of the daily requirement of cholesterol covered by that food.
13. Sodium - the amount of sodium in that food. (mg)
14. Sodium (% Daily Value) - percentage of the daily requirement of sodium covered by that food.
15. Carbohydrates - the amount of carbohydrates in that food. (grams)
16. Carbohydrates (% Daily Value) - percentage of the daily requirement of carbohydrates covered by that food.
17. Dietary Fiber - the amount of dietary fiber in that food. (grams)
18. Dietary Fiber (% Daily Value) - percentage of the daily requirement of dietary fiber covered by that food.
19. Sugars - the amount of sugars in that food. (grams)
20. Protein - the amount of protein in that food. (grams)
21. Vitamin A (% Daily Value) - percentage of the daily requirement of vitamin A covered by that food.
22. Vitamin C (% Daily Value) - percentage of the daily requirement of vitamin C covered by that food.
23. Calcium (% Daily Value) - percentage of the daily requirement of calcium covered by that food.
24. Iron (% Daily Value) - percentage of the daily requirement of iron covered by that food.

- Exploring data

1. Category - Breakfast, Beef & Pork, Chicken & Fish, Salads, Snacks & Sides, Desserts, Beverages, Coffee & Tea, Smoothies & Shakes.
2. Serving Size - From 9.6 oz (272 g) to 1 carton (236 ml)
3. Calories - From 1880 kcal to 0 kcal.

4. Calories from Fat - From 1060 kcal to 0 kcal.
5. Total fat - From 118g to 0g.
6. Total Fat (% of Daily Value) - From 182% to 0%.
7. Saturated Fat - From 20g to 0g.
8. Saturated Fat (% Daily Value) - From 102% to 0%.
9. Trans Fat - From 2.5g to 0g.
10. Cholesterol - From 575mg to 0mg.
11. Cholesterol (% Daily Value) - From 192% to 0%.
12. Sodium - From 3600mg to 0mg.
13. Sodium (% Daily Value) - From 150% to 0%.
14. Carbohydrates - From 141g to 0g.
15. Carbohydrates (% Daily Value) - From 47% to 0%.
16. Dietary Fiber - From 7g to 0g.
17. Dietary Fiber (% Daily Value) - From 28% to 0%.
18. Sugars - From 128g to 0g.
19. Protein - From 87g to 0g.
20. Vitamin A (% Daily Value) - From 170% to 0%.
21. Vitamin C (% Daily Value) - From 240% to 0%.
22. Calcium (% Daily Value) - From 70% to 0%.
23. Iron (% Daily Value) - From 40% to 0%.

Exploring the data, we did not catch any signs of data quality problems.

- **Verifying data quality**

We have access to the data we need and it is good enough to support our goals.

TASK 4. Planning your project

TO-DO list

- Clean data if necessary(null values, transform oz to ml,several weight/volume units transfer em all to grams). 1-2hours
- check data quality 30min
- Describe the data, find avg calories,most calorie dense category 2 hours
- Find nutrition correlation 2 hours
- Find the distribution of food categories in the menu 2 hours
- McDonald's value meal average calories 30 min
- Average daily caloric needs for kids vs calories in McDonald's Kids menu 2-3h
- Explore content of daily average % of cholesterol and sodium in McDonald's menu versus other nutrition 3-5h
- Find the least number of items you could order from the menu to meet one day's nutritional requirements 4-7h
- Find the healthiest combination of items from McDonald's menu to get daily nutritional requirements 3-6h

Methods and tools we plan to use: pandas,numpy,matplotlib,seaborn and probably more on the go.